Annual Report for the Period 5-1-97 to 10-31-97

ONR-N00014-97-1-0509
Office of Naval Research
Mathematical, Computer, and Information Sciences Division

Structure and Geometry in Partial Differential Equations for Image Processing and Analysis

Guillermo Sapiro
Department of Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455, USA
guille@ee.umn.edu
Phone: (612) 625-1343
FAX: (612) 625-4583

October 20, 1997

Contents

Part I
Introduction

In this document we describe work that is being performed under the support of the grant ONR-N00014-97-1-0509, from the Mathematical, Computer, and Information Sciences Division, Office of Naval Research.

We first describe results on shape-base image contrast enhancement. This work shows how to perform local contrast enhancement while preserving the shapes in the image. We have transfered this software to Dr. Hewer and more recently to Dr. Szymczak from Physical Acoustics at the Naval Research Laboratory for testing on underwater laser images (LIDAR). We expect further collaborations with ONR in this area.

We then show a novel approach to anisotropic diffusion. This approach is based on robust statistics, and in particular, in the theory of influence functions. This technique is, for example, of particular significance for image denoising prior to segmentation.

We conclude the technical part of this document with a description of a novel technique of incorporating prior information in anisotropic diffusion. The idea is to use Bayes rule to compute posterior distributions, and then, regularize those distributions via partial differential equations before the MAP is computed.

Although a number of results have already been obtained in these areas, see Part V, the work described in this document is still in progress. As was mentioned above, we want to extend the work on shape-preserving contrast enhancement to include additional definitions of shape and adapt it to specific applications. We are also planing to extend the robust framework for anisotropic diffusion to vector-valued data, and investigate fast implementations, inspired by the work of Chan and colleagues [8]. We plan to further investigate the underlying theory of the posterior diffusion work, and to apply it to additional problems. The work described in this document opens a number of theoretical questions that we plan to address as well.

We should also note that being this the first period of this grant, and my first months at the University of Minnesota, part of the ONR support was used to buy equipment that is being used to build an image processing and computer vision laboratory.

Part II
Shape Preserving Local Contrast Enhancement

1 Introduction

Images are captured at low contrast in a number of different scenarios. The main reason for this problem is poor lighting conditions (e.g., pictures taken at night or against the sun rays).

As a result, the image is too dark or too bright, and is inappropriate for visual inspection or simple observation. The most common way to improve the contrast of an image is to modify its pixel value distribution, or *histogram.* A schematic example of the contrast enhancement problem and its solution via histogram modification is given in Figure 1. On the left, we see a low contrast image with two different squares, one inside the other, and its corresponding histogram. We can observe that the image has low contrast, and the different objects can not be identified, since the two regions have almost identical grey values. On the right we see what happens when we modify the histogram in such a way that the grey values corresponding to the two regions are separated. The contrast is improved immediately.

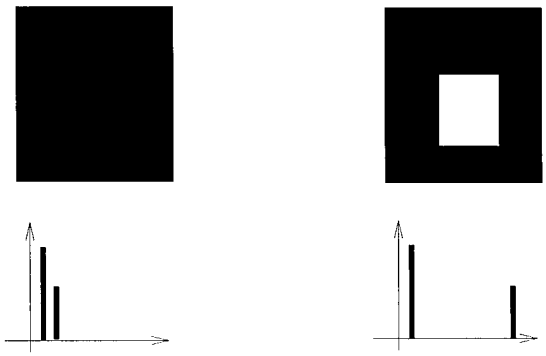


Figure 1: *Schematic explanation of the use of histogram modification to improve image contrast.*

Histogram modification, and in particular histogram equalization (uniform distributions), is one of the basic and most useful operations in image processing, and its description can be found in any book on image processing. This operation is a particular cases of homomorphic transformations: Let $\Omega \subseteq I\!R^2$ be the image domain and $u : \Omega \to [a, b]$ the given (low contrast) image. Let $h : [a, b] \to [c, d]$ be a given function which we assume to be increasing. The image $v := h(u)$ is called an homomorphic transformation of u. The particular case of histogram equalization corresponds to selecting h to be the *distribution function* H of u:

$$H(\lambda) := \frac{Area\{x \in \Omega : u(x) \leq \lambda\}}{Area(\Omega)} \tag{1}$$

If we assume that H is strictly increasing, then the change of variables

$$v(x) = (b - a)H(u(x)) + a, \tag{2}$$

gives a new image whose distribution function is uniform in the interval $[a, b]$, $a, b \in I\!R$, $a < b$. This useful and basic operation has an important property which, in spite of being obvious, we would like to acknowledge: neither it creates or destroys image information.

As argued by the *Mathematical Morphology* school [1, 27], the basic operations on images should be invariant with respect to contrast changes, i.e., homomorphic transformations. As

a consequence, it follows that the basic information of an image is contained in the family of its *binary shadows* or *level-sets*, that is, in the family of sets

$$X_\lambda u := \{x \in \Omega : u(x) \geq \lambda\}, \tag{3}$$

for all values of λ in the range of u. Observe that, under fairly general conditions, an image can be reconstructed from its level-sets by the formula $u(x) = sup\{\lambda : x \in X_\lambda u\}$. If h is a strictly increasing function, the transformation $v = h(u)$ does not modify the family of level-sets of u, it only changes its index in the sense that $X_{h(\lambda)}v = X_\lambda u$ for all λ.

Although one can argue if *all* operations in image processing must hold this principle, for the purposes of the present work we shall stick here to this basic principle. There are a number of reasons for this. First of all, a considerable large amount of the research in image processing is based on assuming that regions with (almost) equal grey-values, which are topologically connected (see below), belong to the same physical object in the 3D world. Following this, it is natural to assume then that the "shapes" in an given image are represented by its level-sets (we will later see how we deal with noise that produces deviations from the level-sets). Furthermore, this commonly assumed image processing principle will permit to develop a theoretical and practical framework for shape preserving contrast enhancement. This can be extended to other definitions of shape, different from the level-sets morphological approach here assumed. We should note that level-set theory is also applicable to a large number of problems beyond image processing [21, 30]. We plan in the future to investigate other definitions of shape for contrast enhancement.

In this work, we want to design *local* histogram modification operations which preserve the family of level-sets of the image, that is, following the morphology school, preserve shape (see [6] for details). Local contrast enhancement is mainly used to further improve the image contrast and facilitate the visual inspection of the data. As we will see later, global histogram modification not always produces good contrast, and specially small regions, are hardly visible after such a global operation. On the other hand, local histogram modification improves the contrast of small regions as well, but since the level-sets are not preserved, artificial objects are created. The theory developed will enjoy the best of both words: The shape-preservation property of global techniques and the contrast improvement quality of local ones.

It is not the goal of this document to review the extensive research done in contrast enhancement in the past. We should only note that basically, histogram modification techniques are divided in the two groups mentioned above, local and global, and their most popular representatives can be found in any basic book in image processing and computer vision. An early attempt to introduce shape criteria in contrast enhancement was done in [10]. To the best of our knowledge, non of the variations to histogram modification reported in the literature have formally approached the problem of shape preserving contrast enhancement as done in our work.

2 A variational formulation

We call *representatives* of u all images of the form $v = h(u)$, where h is a strictly increasing function. The question is which representative of u is the best for our purposes. That will

depend, of course, in what our purposes are. We have seen above which is the function h we have to take if we want to normalize the contrast making the distribution function of u uniform. In addition, it was shown in [28] that when equalizing an image $u : \Omega \to [a, b]$ on the range $[a, b]$ we are minimizing the functional

$$E(v) = \frac{|\Omega|}{2(b-a)} \int_\Omega \left(v(x) - \frac{b-a}{2} \right)^2 dx - \frac{1}{4} \int_\Omega \int_\Omega |v(x) - v(z)| dx dz.$$

The second term of the integral can be understood as a measure of the contrast of the whole image. Thus when minimizing $E(v)$ we are distributing the values of u so that we maximize the contrast. The first term tries to keep the values of u as near as possible to the mean $(b-a)/2$. When minimizing E on the class of functions with the same family of binary shadows as u we get the equalization of u. We will see below how to modify this energy to obtain shape preserving local contrast enhancement.

3 Connected components

To be able to extend the global approach to a local setting we have to insist in our main constraint: We have to keep the same topographic map, that is, we have to keep the same family of level-sets of u but we have the freedom to assign them a "convenient" grey level. To make this statement more precise, let us give some definitions (see [31]).

Definition 1 *Let X be a topological space. We say that X is connected if it cannot be written as the union of two nonempty closed (open) disjoint sets. A subset C of X is called a connected component if C is a maximal connected subset of X, i.e., C is connected and for any connected subset C_1 of X such that $C \subseteq C_1$, then $C_1 = C$.*

This definition will be applied to subsets X of $\mathbb{R}^2$ which are topological spaces with the topology induced from $\mathbb{R}^2$, i.e., an open set of X is the intersection of an open set of $\mathbb{R}^2$ with X. We shall need the following observation which follows from the definition above: Two connected components of a topological space are either disjoint or they coincide; thus the topological space can be considered as the disjoint union of its connected components.

There are several notions of connectivity for a topological space. One of the most intuitive ones is the notion of arcwise connected (also called connected by arcs). A topological space X is said to be connected by arcs if any two points x, y of X can be joined by an arc, i.e., there exists a continuous function $\gamma : [0, 1] \to X$ such that $\gamma(0) = x, \gamma(1) = y$. In a similar way as above we define the connected components (with respect to this notion of connectivity) as the maximal connected set. These notions could be used below instead of the one given in Definition 1.

Definition 2 *Let $u : \Omega \to [a, b]$ be a given image. A section of the topographic map of u is a set of the form*

$$X_{\lambda_1, \lambda_2} = \cup_{\lambda \in [\lambda_1, \lambda_2]} C_\lambda, \tag{4}$$

where C_λ is a connected component of $[u = \lambda]$ such that for each $\lambda', \lambda'' \in [\lambda_1, \lambda_2]$, $\lambda' < \lambda''$, the set

$$X_{\lambda',\lambda''} = \cup_{\lambda\in[\lambda',\lambda'']} C_\lambda \tag{5}$$

is also connected.

Definition 3 *Let $u : \Omega \to [a, b]$ be a given image and let $\{X_\lambda : \lambda \in [a, b]\}$ be the family of its level-sets. We shall say that the mapping $h : \Omega \times I\!R \to I\!R$ is a local contrast change if the following properties hold:*

P1: *h is continuous in the following sense: $h(z, \lambda') \to h(x, \lambda)$ when $z \to x, \lambda' \to \lambda, z \in X_{\lambda'}, x \in C_\lambda$, C_λ being a connected component of $[u = \lambda]$.*

P2: *$h(x, \cdot)$ is an increasing function of λ for all $x \in \Omega$.*

P3: *$h(x, \lambda) = h(y, \lambda)$ for all x, y are in the same connected component of $[u = \lambda]$, $\lambda \in I\!R$.*

P4: *Let Γ be a connected set with $u(\Gamma)$ not reduced to a point. Let $v(x) = h(x, u(x))$. Then $v(\Gamma)$ is not reduced to a point.*

P5: *Let $X_{\lambda_1,\lambda_2} = \cup_{\lambda\in[\lambda_1,\lambda_2]} C_\lambda$ be a section of the topographic map of u and let $x \in C_{\lambda_1}, y \in C_{\lambda_2}$. Then $h(x, \lambda_1) < h(y, \lambda_2)$.*

Definition 4 *Let $u : \Omega \to [a, b]$ be a given image. We shall say that v is a local representative of u if there exists some local contrast change h such that $v(x) = h(x, u(x))$, $x \in \Omega$.*

We collect in the next proposition some properties which follow immediately from the definitions above.

Proposition 1 ([6]) *Let $u : \Omega \to [a, b]$ and let $v(x) = h(x, u(x))$, $x \in \Omega$, be a local representative of u. Then*

1. *$v(x) = sup\{h(x, \lambda) : x \in X_\lambda u, x \in \Omega\}$. We have that $x \in X_\lambda u$ if and only if $x \in X_{h(x,\lambda)} v$, $x \in \Omega$, $\lambda \in I\!R$.*
2. *v is a continuous function.*
3. *Let Γ (Γ') be a connected component of $[v = \mu]$ (resp. $[u = \lambda]$) containing x, $\mu = h(x, \lambda)$. Then $\Gamma = \Gamma'$.*
4. *Let X_{λ_1,λ_2} be a section of the topographic map of u. Then X_{λ_1,λ_2} is also a section of the topographic map of v.*

The previous proposition can be phrased as saying that the set of "objects" contained in u is the same as the set of "objects" contained in v, if we understand the "objects" of u as the connected connected components of the level-sets $[\lambda \leq u < \mu]$, $\lambda < \mu$, and respectively for v.

Remark: Our definition of local representative is contained in the notion of *dilation* as given in [27], Theorem 9.3. Let $\mathcal{U}_n$ be a lattice of functions $f : I\!R^n \to I\!R^n$. A mapping $\Gamma : \mathcal{U}_n \to \mathcal{U}_n$ is called a dilation of $\mathcal{U}_n$ if and only if it can be written as

$$\Gamma(f)(x) = sup\{g(x; y, t) : y \in I\!R^n, t \leq f(y)\}, \quad x \in I\!R^n,$$

where $g(x; y, t)$ is a function assigned to each point $(y, t) \in I\!R^n \times I\!R$ and is possibly different from point to point. Thus, let h be a local contrast change and let $v(x) = h(x, u(x))$. Let us denote by $X_t(f, x)$ the connected component of $X_t f$ which contains x if $x \in X_t f$, otherwise, let $X_t(f, x) = \emptyset$. Let $g(x; y, t) := h(x, t)$ if $X_t(f, x) \cap X_t(f, y) \neq \emptyset$; and $:= 0$ if $X_t(f, x) \cap X_t(f, y) = \emptyset$. Then $v = \Gamma(u)$.

4 Shape preserving enhancement

We can now state precisely the main question we want to address: *what is the best local representative v of u*, when the goal is to perform local contrast enhancement while preserving the connected components (and level-sets). For that we shall use the energy formulation given above. Let A be a connected component of the set $[\lambda \leq u < \mu]$, $\lambda, \mu \in I\!R$, $\lambda < \mu$. Write

$$E(v, A) := \frac{|A|}{2(\mu - \lambda)} \int_A \left(v(x) - \frac{\mu - \lambda}{2} \right)^2 dx - \frac{1}{4} \int_A \int_A |v(x) - v(z)| dx dz. \tag{6}$$

We then look for a local representative v of u which minimizes $E(v, A)$ for all connected components A of all sets of the form $[\lambda \leq u < \mu]$, $\lambda, \mu \in I\!R$, $\lambda < \mu$, or, in other words, the distribution function of v in all connected components of $[\lambda \leq v < \mu]$ is uniform in the range $[\lambda, \mu]$, for all $\lambda, \mu \in I\!R$, $\lambda < \mu$. We now show how to solve this problem.

Let us introduce some notation that will make our discussion easier. Without loss of generality we assume that $u : \Omega \to [0, 1]$. Let $\lambda_{k,j} := j/2^k$, $k = 0, 1, 2, ...$, $j = 0, ..., 2^k$. We need to assume that H, the distribution function of u, is continuous and strictly increasing. For that we assume that u is continuous and

$$Area\{x \in \Omega : u(x) = \lambda\} = 0, \quad \text{for all } \lambda \in I\!R. \tag{7}$$

We shall construct a sequence of functions converging to the solution of the problem. Let $w_0 = H(u)$ be the histogram equalization of u. Suppose that we already constructed $w_0, ..., w_{i-1}$. Let us construct w_i. For each $j = 0, 1, ..., 2^i - 1$, let

$$O_{i,j} := [\lambda_{i,j} \leq w_{i-1} < \lambda_{i,j+1}], \tag{8}$$

and let $O_{i,j;r}$ be the connected components of $O_{i,j}$, $r = 1, ..., n_{i,j}$ ($n_{i,j}$ can be eventually ∞). Define

$$h_{i,j;r}(\lambda) := \frac{|[w_{i-1} \leq \lambda] \cap O_{i,j;r}|}{|O_{i,j;r}|} (\lambda_{i,j+1} - \lambda_{i,j}) + \lambda_{i,j} \ . \tag{9}$$

By our assumption (7), $h_{i,j;r}$ is a continuous strictly increasing function and we can equalize the histogram of w_{i-1} in $O_{i,j;r}$. Thus, we define

$$w_{i,j;r} := h_{i,j;r}(w_{i-1})\chi_{O_{i,j;r}}, \tag{10}$$

$j = 0, 1, ..., 2^i - 1$,$r = 1, ..., n_{i,j}$, and

$$w_i := \sum_{j=1}^{2^i-1} \sum_{r=1}^{n_{i,j}} w_{i,j;r}\chi_{O_{i,j;r}}. \tag{11}$$

We then have the following results:

Theorem 1 ([6]) *Under the assumption* (7) *the functions w_i have a uniform histogram for all connected components of all "dyadic" sets of the form $[\lambda \leq w_i < \mu]$ where $\lambda, \mu \in \{\lambda_{i,j} : j = 0, ..., 2^i\}$, $\lambda < \mu$. Moreover, as $i \to \infty$, w_i converges to a function w which has a uniform histogram for all connected components of all sets $[\lambda \leq w < \mu]$, for all $\lambda, \mu \in [0, 1]$, $\lambda < \mu$.*

Theorem 2 ([6]) *Let w be the function constructed in Theorem* 1. *Then w is a local representative of u.*

5 The algorithm and examples

The algorithm has been described in the previous section. Let us summarize it here. Let $u : \Omega \to [0, M]$ be an image whose values have been normalized in $[0, M]$. Let $\lambda_{k,j} := jM/2^k$, $k = 0, 1, 2, .., N$, $j = 0, ..., 2^k$
Step 1: Construct $w_0 = H(u)$ be the histogram equalization of u.
Step 2: Construction of w_i, $i = 1, ..., N$.
Suppose that we already constructed $w_0, ..., w_{i-1}$. Let us construct w_i. For each $j = 0, 1, ..., 2^i - 1$, let

$$O_{i,j} := [\lambda_{i,j} \leq w_{i-1} < \lambda_{i,j+1}], \tag{12}$$

and let $O_{i,j;r}$ be the connected components of $O_{i,j}$, $r = 1, ..., n_{i,j}$. Let $h_{i,j;r}$ be the distribution function of $w_{i-1}\chi_{O_{i,j;r}}$ with values in the range $[\lambda_{i,j}, \lambda_{i,j+1}]$.

Then we define

$$w_i := \sum_{j=1}^{2^i-1} \sum_{r=1}^{n_{i,j}} h_{i,j;r}(w_{i-1})\chi_{O_{i,j;r}}. \tag{13}$$

Then we equalize w_0 in all connected components of $O_{1,0}$ in the range $[0, m_{0,1} - 1]$, respectively in all connected components of $O_{1,1}$ in the range $[m_{0,1}, M]$. In this way we construct w_1. Then we compute the mean values of w_1 in $O_{1,0}$, $O_{1,1}$. Denote them by $m_{1,1}$, $m_{1,3}$ ($m_{1,2} = m_{0,1}$). Now we use these values to subdivide again w_1 into four pieces and proceed to equalize the histogram of w_1 in all connected components of all these pieces. We may continue iteratively in this way until desired.

Figure 2 illustrates the importance of performing shape preserving local contrast enhancement. Fig. 2b shows the global histogram equalization of Fig. 2a. Fig. 2c shows the result of classical local histogram equalization [15]. Fig. 2d presents the result of our algorithm applied to Fig. 2a. The level-lines off all the figures are given in Fig. 2e-2h respectively. We see how different connected components do not interact in the proposed scheme, and the contrast is improved while preserving the objects in the scene.

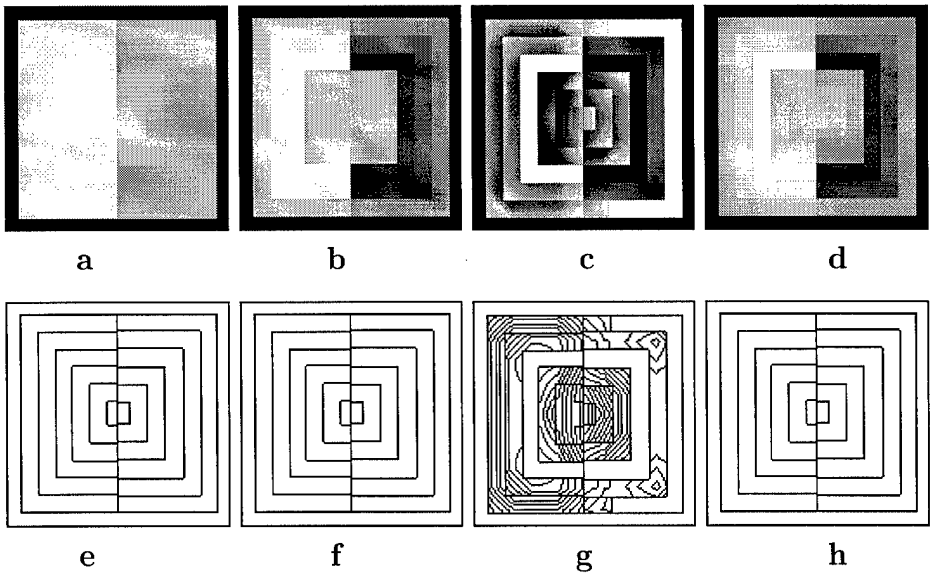


Figure 2: *Example of the level-sets preservation. The first row show the original image, global histogram modification, classical local modification, and the proposed shape preserving local histogram modification. The second row shows the corresponding level-sets.*

Results for a real image are presented in Figure 3. Fig. 3b show the global equalization of Fig. 3a. Fig. 3c shows an intermediate step of the proposed algorithm, while Fig. 3d is the steady state solution. Note how objects that are not visible in the global modification, like those trough the window, are now visible with the new local scheme.

Experiments with a color image are given in Figure 4, working on the YIQ color space. In Fig. 4a we present the original image. In Fig. 4b, we apply the proposed local histogram modification to the Y channel only, re-scaling the chrominance vector to maintain the same color point on the Maxwell triangle.

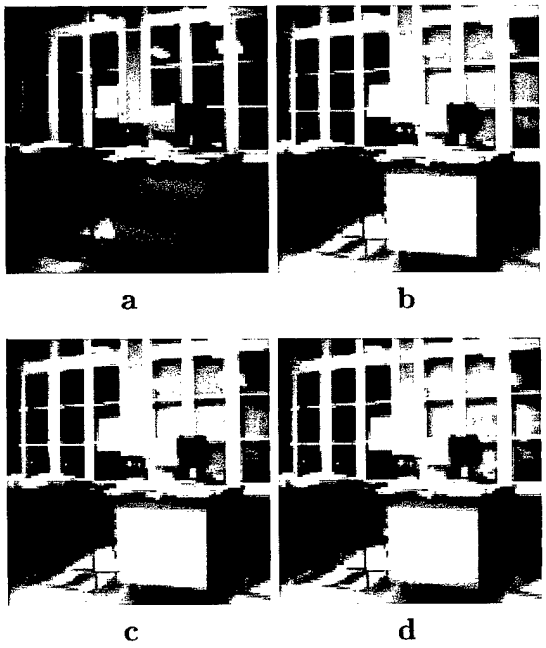


Figure 3: *Example of shape preserving local histogram modification for real data. The first row shows the original image (a) and the result of global histogram modification (b). An intermediate state (c), together with the steady state of the proposed algorithm (d) are shown in the second row.*

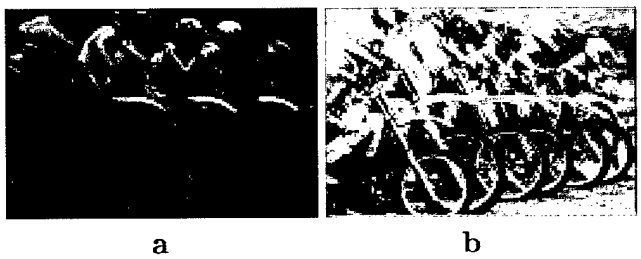


Figure 4: *Example of local histogram modification of a color image. (a) Original image. (b) The proposed algorithm is applied only to the Y channel, re-scaling the chrominance vector to maintain the same color point on the Maxwell triangle. (This is a color figure.)*

Part III
Influence-based anisotropic diffusion

6 Introduction

Since the elegant formulation of anisotropic diffusion introduced by Perona and Malik [22], a considerable amount of research has been devoted to the theoretical and practical understanding of this and related methods for image enhancement. See [2, 4, 23] and references therein. We develop a statistical interpretation of anisotropic diffusion, specifically, from the point of view of robust statistics. We show that the Perona-Malik [22] diffusion equation is equivalent to a robust procedure that estimates a piecewise constant image from a noisy input image. The "edge-stopping" function in the anisotropic diffusion equation is closely related to the error norm and influence function in the robust estimation framework. We exploit this robust statistical interpretation of anisotropic diffusion to choose alternative robust error norms, and hence, alternative "edge-stopping" functions. In particular, we propose a new "edge-stopping" function based on *Tukey's biweight* robust error norm, which preserves sharper boundaries than previous formulations and improves the automatic stopping of the diffusion. The robust statistical interpretation also provides a means for detecting the boundaries (edges) between the piecewise constant regions. These boundaries are considered to be "outliers" in the robust estimation framework. Edges in a smoothed image are, therefore, very simply detected as those points that are treated as outliers. Details, examples, and extensions, including connections to line processing and techniques as those described in [9, 13], can be found in [4]. We also propose, following [29], possible extensions of this method to color images and vector data in general, which will be further investigated, and describe the use of the theory here presented for image sharpening.

7 Anisotropic diffusion

Diffusion algorithms remove noise from an image by modifying the image via a partial differential equation (PDE). For example, consider applying the isotropic diffusion equation (the heat equation) given by $\frac{\partial I(x,y,t)}{\partial t} = \mathrm{div}(\nabla I)$, using the original (degraded/noisy) image $I(x, y, 0)$ as the initial condition, where $I(x, y, 0) : I\!R^2 \rightarrow I\!R^+$ is an image in the continuous domain, (x, y) specifies spatial position, t is an artificial time parameter, and where ∇I is the image gradient.

Perona and Malik [22] replaced the classical isotropic diffusion equation with

$$\frac{\partial I(x,y,t)}{\partial t} = \mathrm{div}(g(\| \nabla I \|)\nabla I), \tag{14}$$

where $\| \nabla I \|$ is the gradient magnitude, and $g(\| \nabla I \|)$ is an "edge-stopping" function. This function is chosen to satisfy $g(x) \rightarrow 0$ when $x \rightarrow \infty$ so that the diffusion is "stopped" across edges.

Perona and Malik discretized their anisotropic diffusion equation as follows:

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(\nabla I_{s,p}) \nabla I_{s,p}, \tag{15}$$

where I_s^t is a discretely-sampled image, s denotes the pixel position in a discrete, two-dimensional grid, and t now denotes discrete time steps (iterations). The constant $\lambda \in I\!R^+$ is a scalar that determines the rate of diffusion, η_s represents the spatial neighborhood of pixel s, and $|\eta_s|$ is the number of neighbors. They linearly approximated the image gradient (magnitude) in a particular direction as $\nabla I_{s,p} = I_p - I_s^t, \;\; p \in \eta_s$

Qualitatively, the effect of anisotropic diffusion is to smooth the original image while preserving brightness discontinuities. As we will see, the choice of $g(\cdot)$ can greatly affect the extent to which discontinuities are preserved.

8 Robust estimation

Our goal is to develop a statistical interpretation of the Perona-Malik anisotropic diffusion equation. Toward that end, we adopt an oversimplified statistical model of an image. In particular, we assume that a given input image is a piecewise constant function that has been corrupted by zero-mean Gaussian noise with small variance.

Consider the image intensity differences, $I_p - I_s$, between pixel s and its neighboring pixels p. Within one of the piecewise constant image regions, these neighbor differences will be small, zero-mean, and normally distributed. Hence, an optimal estimator for the "true" value of the image intensity I_s at pixel s minimizes the square of the neighbor differences. This is equivalent to choosing I_s to be the mean of the neighboring intensity values.

The neighbor differences will not be normally distributed, however, for an image region that includes a boundary (intensity discontinuity). The intensity values of the neighbors of an edge pixel s are drawn from two different populations, and in estimating the "true" intensity value at s we want to include only those neighbors that belong to the same population. With respect to our assumption of Gaussian noise within each constant region, if p and s belong to two different sides of the edge, the neighbor difference $I_p - I_s$ can be viewed as an *outlier* because it does not conform to the statistical assumptions.

The field of robust statistics [16, 17] is concerned precisely with estimation problems in which the data contains gross errors, or outliers. See for example [4] for references to the applications of robust statistics to image processing and computer vision.

Motivated then by robust statistics, we wish to find an image I that satisfies the following optimization criterion:

$$\min_I \sum_{s \in I} \sum_{p \in \eta_s} \rho(I_p - I_s, \sigma), \tag{16}$$

where $\rho(\cdot)$ is a robust error norm and σ is a "scale" parameter that will be discussed further below. To minimize (16), the intensity at each pixel must be "close" to those of its neighbors. As we shall see, an appropriate choice of the ρ-function allows us to minimize the effect of the outliers at the boundaries between piecewise constant image regions.

Equation (16) can be solved by gradient descent:

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} \psi(I_p - I_s^t, \sigma), \tag{17}$$

where $\psi(\cdot) = \rho'(\cdot)$, and t again denotes the iteration. The update is carried out simultaneously at every pixel s.

The specific choice of the robust error norm or ρ-function in (16) is critical. To analyze the behavior of a given ρ-function, we consider its derivative ψ, which is proportional to the *influence function* [16]. This function characterizes the bias that a particular measurement has on the solution. For example, the quadratic ρ-function has a linear ψ-function.

If the distribution of values $(I_p - I_s^t)$ in every neighborhood is a zero-mean Gaussian, then $\rho(x, \sigma) = x^2/\sigma^2$ provides an optimal local estimate of I_s^t. This *least-squares* estimate of I_s^t is, however, very sensitive to outliers because the influence function increases linearly and without bound. For a quadratic ρ, I_s^{t+1} is assigned to be the mean of the neighboring intensity values I_p. When these values come from different populations (across a boundary) the mean is not representative of either population, and the image is blurred too much. Hence, the quadratic gives outliers (large values of $|\nabla I_{s,p}|$) too much *influence.*

To increase robustness and *reject* outliers, the ρ-function must be more forgiving about outliers; that is, it should increase less rapidly than x^2. For example, consider the *Lorentzian* error norm (see Figure 5):

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \qquad \psi(x, \sigma) = \frac{2x}{2\sigma^2 + x^2}. \tag{18}$$

Examination of the ψ-function reveals that, when the absolute value of the gradient magnitude increases beyond a fixed point determined by the scale parameter σ, its influence is reduced. We refer to this as a *redescending* influence function [16].

9 Robust statistics and anisotropic diffusion

We now explore the relationship between robust statistics and anisotropic diffusion by showing how to convert back and forth between the formulations. Recall the continuous anisotropic diffusion equation (14). The continuous form of the robust estimation problem in (16) can be posed as:

$$\min_I \int_\Omega \rho(\| \nabla I \|) d\Omega, \tag{19}$$

where Ω is the domain of the image and where we have omitted σ for notational convenience. One way to minimize (19) is via gradient descent using the calculus of variations:

$$\frac{\partial I(x, y, t)}{\partial t} = \operatorname{div}\left(\rho'(\| \nabla I \|) \frac{\nabla I}{\| \nabla I \|}\right). \tag{20}$$

By defining $g(x) \doteq \frac{\rho'(x)}{x}$, we obtain the straightforward relation between image reconstruction via robust estimation (19) and image reconstruction via anisotropic diffusion (14). (See for example [22, 34] for previous uses of this relation.)

The same relationship holds for the discrete formulation; compare (15) and (17) with $\psi(x) = \rho'(x) = g(x)\,x$. Note that additional terms will appear in the gradient descent equation if the magnitude of the image gradient is discretized in a nonlinear fashion. In the remainder of this part of the document we proceed with the discrete formulation as given in previous section. The basic results we present hold for the continuous domain as well.

Perona and Malik suggested two different edge stopping functions ($g(\cdot)$) in their anisotropic diffusion equation. Each of these can be viewed in the robust statistical framework by converting the $g(\cdot)$ functions into the related ρ-functions. They first suggested $g(x) = \frac{1}{1+\frac{x^2}{K^2}}$, for a positive constant K. It is easy to see, [4, 19, 34], that this edge stopping function corresponds to the Lorentzian norm of robust statistics. The other g-function proposed by Perona and Malik is related to the robust error norm proposed by Leclerc.

10 Exploiting the relationship

The above derivations demonstrate that anisotropic diffusion is the gradient descent of an estimation problem with a familiar robust error norm. What's the advantage of knowing this connection? We answer this question now.

While the Lorentzian is more robust than the quadratic norm, its influence does not descend all the way to zero. We can choose a more "robust" norm from the robust statistics literature which does descend to zero, as the *Tukey's biweight*, whose influence function is plotted in Figure 5:

$$\rho(x,\sigma) = \begin{cases} \frac{x^2}{\sigma^2} - \frac{x^4}{\sigma^4} + \frac{x^6}{3\sigma^6} & |x| \le \sigma \\ \frac{1}{3} & \text{otherwise} \end{cases} \tag{21}$$

$$\psi(x,\sigma) = \begin{cases} x(1-(x/\sigma)^2)^2 & |x| \le \sigma, \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

Another error norm from the robust statistics literature, is Huber's *minimax* norm [17] (see also [26, 34]), whose influence function is also plotted in Figure 5:

$$\rho(x,\sigma) = \begin{cases} x^2/2\sigma + \sigma/2 & |x| \le \sigma, \\ |x| & |x| > \sigma, \end{cases} \tag{23}$$

$$\psi(x,\sigma) = \begin{cases} x/\sigma, & |x| \le \sigma, \\ \text{sign}(x) & |x| > \sigma. \end{cases} \tag{24}$$

We would like to compare the influence (ψ-function) of these three norms, but a direct comparison requires that we dilate and scale the functions to make them as similar as possible.

First, we need to determine how large the image gradient can be before we consider it to be an outlier. We appeal to tools from robust statistics to automatically estimate the "robust variance," σ_e, of the image as the MAD [25]:

$$\sigma_e = 1.4826\ \text{median}_I(\|\ \nabla I - \text{median}_I(\|\ \nabla I\ \|)\ \|) \tag{25}$$

where the constant is derived from the fact that the MAD of a zero-mean normal distribution with unit variance is $0.6745 = 1/1.4826$. For a discrete image, the robust variance, σ_e, is computed using the gradient magnitude approximation introduced before.

Second, we choose values for the scale parameters σ to dilate each of the three influence functions so that they begin rejecting outliers at the same value: σ_e. The point where the influence of outliers first begins to decrease occurs when the derivative of the ψ-function is zero. For the modified L_1 or Huber's norm this occurs at $\sigma_e = \sigma$. For the Lorentzian norm it occurs at $\sigma_e = \sqrt{2}\sigma$ and for the Tukey norm it occurs at $\sigma_e = \sigma/\sqrt{5}$. Defining σ with respect to σ_e in this way we plot the influence functions for a range of values of x in Figure 5*a*. Note how each function now begins reducing the influence of measurements at the same point.

Third, we scale the three influence functions so that they return values in the same range. To do this we take λ in (15) to be one over the value of $\psi(\sigma_e, \sigma)$. The scaled ψ-functions are plotted in Figure 5*b*.

Now we can compare the three error norms directly. The modified L_1 norm gives all outliers a constant weight of one while the Tukey norm gives *zero* weight to outliers whose magnitude is above a certain value. The Lorentzian (or Perona-Malik) norm is in between the other two. Based on the shape of $\psi(\cdot)$ we would predict that diffusing with the Tukey norm produces sharper boundaries than diffusing with the Lorentzian (standard Perona-Malik) norm, and that both produce sharper boundaries than the modified L_1 norm. We can also see how the choice of function affects the "stopping" behavior of the diffusion; given a piecewise constant image where all discontinuities are above a threshold, the Tukey function will leave the image unchanged whereas the other two functions will not.

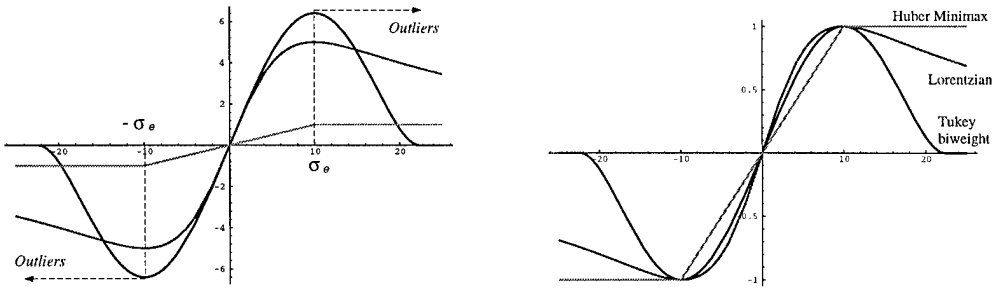


Figure 5: *Lorentzian, Tukey, and Huber ψ-functions. Left: Values of σ chosen as a function of σ_e so that outlier "rejection" begins at the same value for each function. Right: The functions aligned and scaled. (This is a color figure.)*

These predictions are born out experimentally, as can be seen in Figure 6. The figure compares the results of diffusing with the Lorentzian and the Tukey functions. The value of $\sigma_e = 10.278$ was estimated automatically using (25) and the values of σ and λ for each function were defined with respect to σ_e as described above. The figure shows the diffused image after 500 iterations of each method. Observe how the Tukey function results in sharper discontinuities. Note that we can detect edges in the smoothed images very simply

by detecting those points that are treated as outliers by the given ρ-function. Figure 6 shows the outliers (edge points) in each of the images, that is, pixels where $|\nabla I_{s,p}| > \sigma_e$.

It is interesting to note that common robust error norms have frequently been proposed in the literature without mentioning the motivation from robust statistics. For example, Rudin *et al.* [26] proposed a formulation that is equivalent to using the L_1 norm. You *et al.* [34] explored a variety of anisotropic diffusion equations and reported better results for some than for others. In addition to their own explanation for this, their results are predicted, following the development presented here, by the robustness of the various error norms they use.

11 Robust anisotropic sharpening

The basic idea behind image sharpening is to add high frequencies to an image. That is, the sharpened image $\hat{I}$ is obtained from the blurred image I via $\hat{I} = I + H(I)$, where $H(\cdot)$ represents some kind of high-pass filter operation, e.g., the Laplacian. The problem with this approach, denoted as *unsharp masking*, is that it also enhances noise, and not only edges. To solve this problem, in [7] the author proposes to mask $H(I)$ with an edge detector. Since the framework here described is natural to detect edges, it is natural as well to accomplish this task. The basic idea is then to perform anisotropic diffusion on the Image I, robustly detect edges based on outliers, and then mask $H(I)$ using these edges. Examples are given in our reports.

12 Vector-valued images

The extension of the results presented above for vector-valued images follows the framework introduced in [29]. The basic idea is that the gradient direction $\frac{\nabla I}{\|\nabla I\|}$ and the gradient magnitude $\| \nabla I \|$ are replaced by concepts derived from the first fundamental form of the vector image. The direction of maximal change θ_+ ("the gradient direction") of the vector data is given by the eigenvector of this fundamental form corresponding to the maximal eigenvalue λ_+, and the value of the maximal change ("the gradient magnitude") is given by a function of both eigenvalues, that is, $f(\lambda_+, \lambda_-)$. Note that θ_+, λ_+, and λ_- depend on all the components of the vector-valued image.

To extend the robust anisotropic diffusion approach to vector data, we have basically two possibilities. One, [29], if to formulate the problem as the minimization of

$$\int \rho(\lambda_+, \lambda_-) d\Omega,$$

selecting ρ to be the Tukey's robust function. The gradient descent of this variational problem will give a system of coupled anisotropic diffusion equations. The second option if to derive directly the anisotropic equation from (20), and evolve each one of the image components I_i according to

$$\frac{\partial I_i}{\partial t} = \text{div}\left(\psi(\lambda_+, \lambda_-)\begin{pmatrix}\cos\theta_+ \\ \sin\theta_+\end{pmatrix}\right),$$

where ψ is the Tukey's influence function. This topic is currently under investigation (see also [5]).

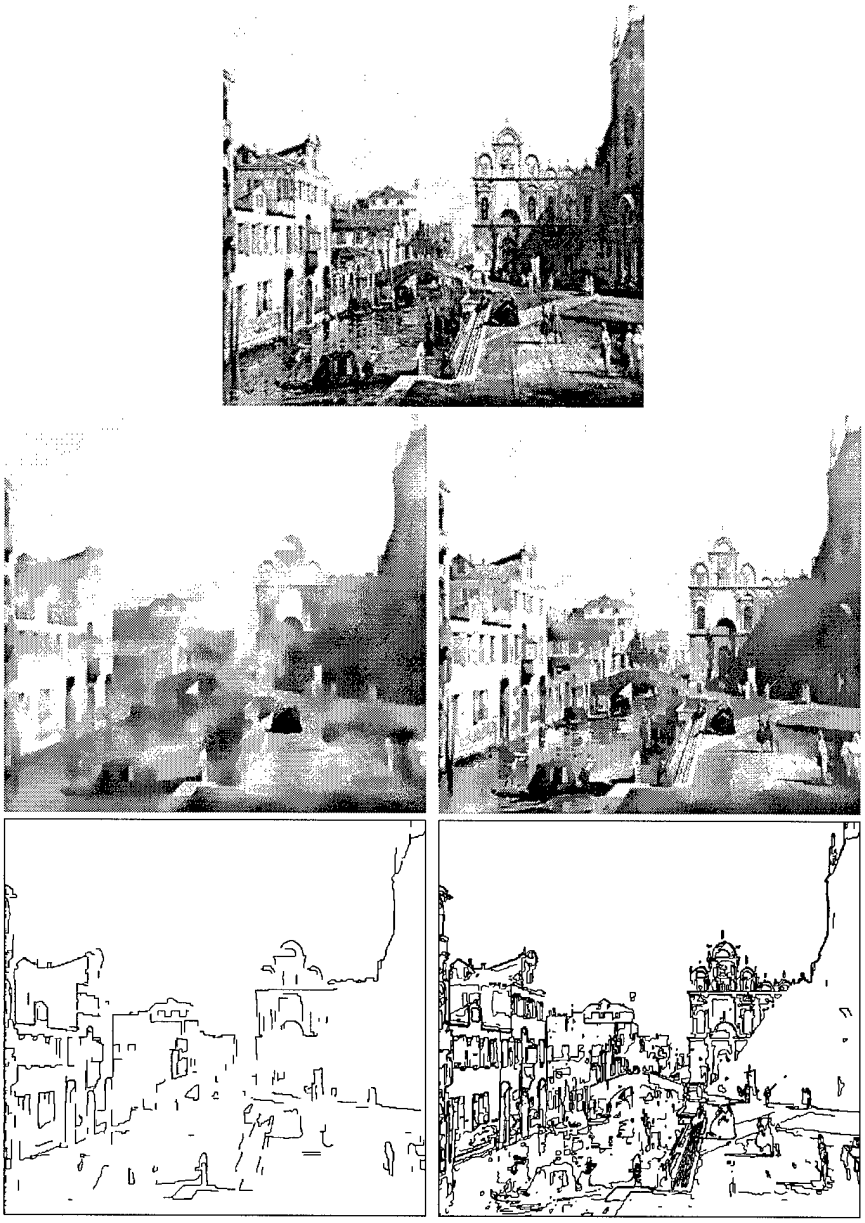


Figure 6: *Comparison of the Perona-Malik (Lorentzian) function (left) and the Tukey function (right) after 500 iterations. The first row shows the original image. The last row shows the edges obtained from the outliers.*

Part IV
Anisotropic Smoothing of Posterior Probabilities

13 Introduction

In [32], we proposed a new segmentation technique that was applied to segmenting MRI volumes of human cortex. The technique comprise three steps. First, the posterior probability of each pixel is computed from its likelihood and a homogeneous prior; i.e., a prior that reflects the relative frequency of each class but is the same across all pixels. Next, the posterior probabilities for each class are anisotropically smoothed (using a 3D-extension of the algorithm suggested by Perona and Malik [22]). Finally, each pixel is classified independently using the MAP rule. Fig. 7 compares the classification of cortical white matter with and without the anisotropic smoothing step. The anisotropic smoothing produces classifications that are qualitatively smoother within regions while preserving detail along region boundaries. The intuition behind the method is straightforward. Anisotropic smoothing of the posterior probabilities results in piecewise constant posterior probabilities which, in turn, yield piecewise "constant" MAP classifications.

We explore the mathematical theory underlying the technique. We demonstrate that prior anisotropic smoothing of the posterior probabilities yields the MAP solution of a discrete MRF with a non-interacting, analog discontinuity field. In contrast, isotropic smoothing of the posterior probabilities is equivalent to computing the MAP solution of a single, discrete MRF using continuous relaxation labeling. Combining a discontinuity field with a discrete MRF is important as it allows the disabling of clique potentials across discontinuities. Furthermore, explicit representation of the discontinuity field suggests new algorithms that incorporate hysteresis and non-maximal suppression. This will be the subject of further investigation.

14 Isotropic smoothing

In this section, we describe the relationship between maximum aposterior probability (MAP) estimation of discrete Markov random fields (MRF) and continuous relaxation labeling (CRL) [24]. This connection was originally made by Li *et. al.* [20]. We review this relationship to introduce the notation that will be used and to point out the similarities between this technique and isotropic smoothing of posterior probabilities. These relationships are depicted in Fig. 8.

We specialize our notation to MRF's defined on image grids. Let $\mathcal{S} = \{1, \ldots, n\}$ be a set of sites where each $s \in \mathcal{S}$ corresponds to a single pixel in the image. For simplicity, we assume that each site can take on labels from a common set $\mathcal{L} = \{1, \ldots, k\}$. Adjacency relationships between sites are encoded by $\mathcal{N} = \{\mathcal{N}_i | i \in \mathcal{S}\}$ where $\mathcal{N}_i$ is the set of sites neighboring site i. Cliques are then defined as subsets of sites so that any pair of sites in a

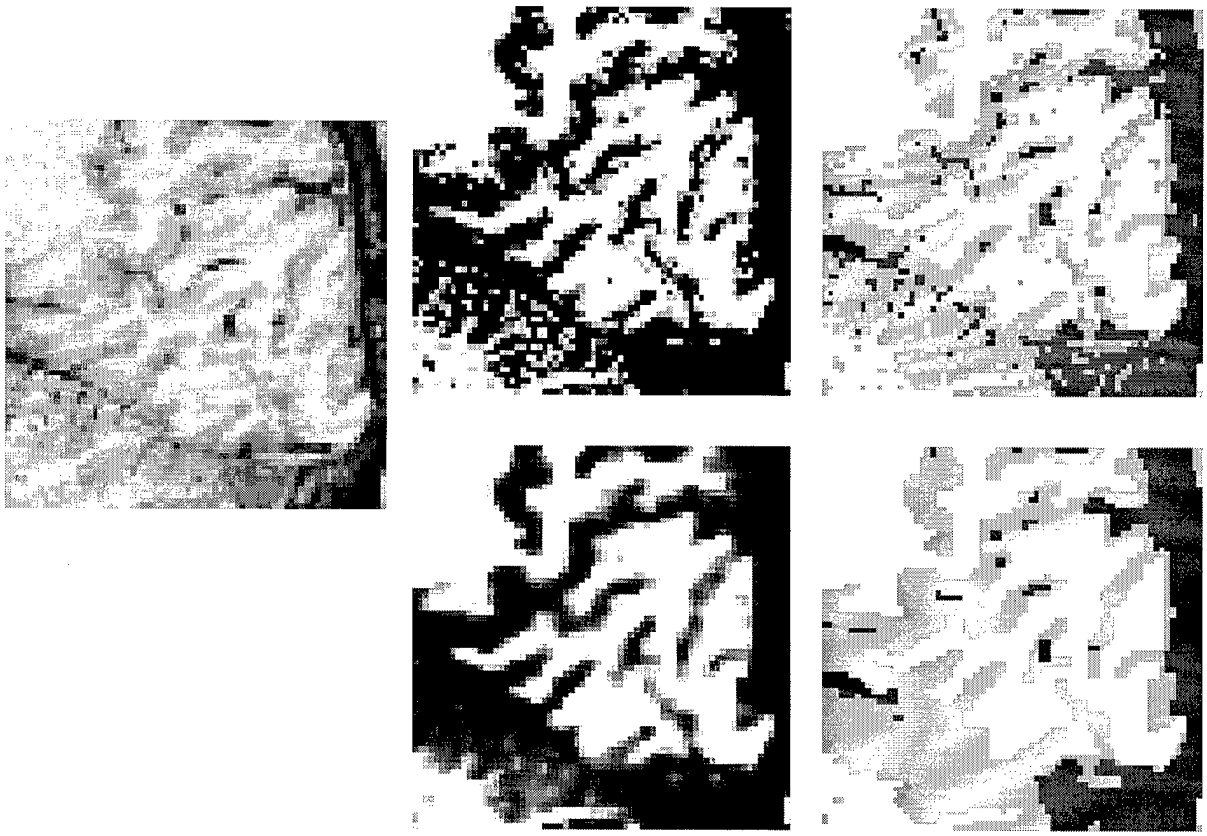


Figure 7: *(Top row) Left: Intensity image of MRI data. Middle: Image of posterior probabilities corresponding to white matter class. Right: Image of corresponding MAP classification. Brighter regions in the posterior image correspond to areas with higher probability. White regions in the classification image correspond to areas classified as white matter; black regions correspond to areas classified as CSF. (Bottom row) Left: Image of white matter posterior probabilities after being anisotropically smoothed. Right: Image of MAP classification computed using smoothed posteriors.*

clique are neighbors. We will only consider 4-neighbor adjacency for images (and 8-neighbor adjacency for volumes) and cliques of sizes no greater than two. By considering each site as a discrete random variable f_i with a probability mass function over $\mathcal{L}$, a discrete MRF $\mathbf{f}$ can be defined over the sites with a Gibbs probability distribution.

If data $d_i \in \mathbf{d}$ is observed at each site i, and is dependent only on its label f_i, then the posterior probability is itself a Gibbs distribution and by the Hammersley-Clifford theorem, also a MRF, albeit a different one [12]: $P(\mathbf{f}|\mathbf{d}) = Z^{-1} \times \exp\{-E(\mathbf{f}|\mathbf{d})\}$ where

$$E(\mathbf{f}|\mathbf{d}) = \sum_{i \in \mathcal{C}_1} V_1(f_i|d_i) + \sum_{(i,j) \in \mathcal{C}_2} V_2(f_i, f_j) \tag{26}$$

and $V_1(f_i|d_i)$ is a combination of the single site clique potential and the independent likelihood. The notation (i, j) refers to a pair of sites; thus, the sum is actually a double sum. Maximizing the posterior probability $P(\mathbf{f}|\mathbf{d})$ is equivalent to minimizing the energy $E(\mathbf{f}|\mathbf{d})$.

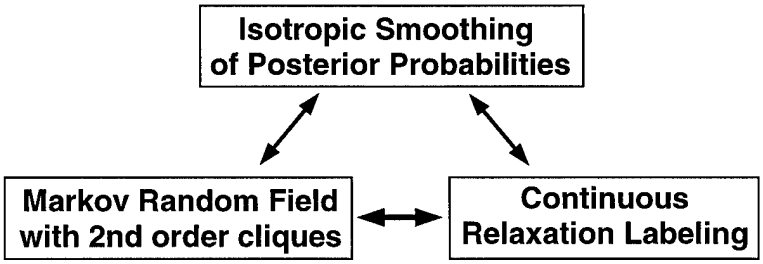


Figure 8: *Equivalence between isotropic smoothing of posterior probabilities, Markov random fields with 2nd order cliques, and continuous relaxation labeling.*

Continuous Relaxation Labeling. The continuous relaxation labeling approach to solving this problem was introduced by Li *et. al.* [20]. In CRL, the class (label) of each site i is represented by a vector $p_i = [p_i(f_i)|f_i \in \mathcal{L}]$ subject to the constraints: (1) $p_i(f_i) \geq 0$ for all $f_i \in \mathcal{L}$, and (2) $\sum_{f_i \in \mathcal{L}} p_i(f_i) = 1$. Within this framework, the energy $E(\mathbf{f}|\mathbf{d})$ to be minimized is rewritten as

$$E(\mathbf{p}|\mathbf{d}) = \sum_{i \in \mathcal{C}_1} \sum_{f_i \in \mathcal{L}} V_1(f_i|d_i) p_i(f_i) + \sum_{(i,j) \in \mathcal{C}_2} \sum_{(f_i, f_j) \in \mathcal{L}^2} V_2(f_i, f_j) p_i(f_i) p_j(f_j). \tag{27}$$

Note that when $p_i(f_i)$ is restricted to $\{0, 1\}$, $E(\mathbf{p}|\mathbf{d})$ reverts to its original counterpart $E(\mathbf{f}|\mathbf{d})$. Hence, CRL embeds the actual combinatorial problem into a larger, continuous, constrained minimization problem.

The constrained minimization problem is typically solved by iterating two steps: (1) gradient computation, and (2) normalization and update. The first step decides the direction that decreases the objective function while the second updates the current estimate while ensuring compliance with the constraints. A review of the normalization techniques that have been proposed are summarized in [20]. Ignoring the need for normalization, continuous relaxation labeling is similar to traditional gradient descent: $p_i^{t+1}(f_i) \leftarrow p_i^t(f_i) - \frac{\partial E(\mathbf{p}|\mathbf{d})}{\partial p_i^t(f_i)}$ where

$$\frac{\partial E(\mathbf{p}|\mathbf{d})}{\partial p_i^t(f_i)} \doteq V_1(f_i|d_i) + 2 \sum_{j:(i,j) \in \mathcal{C}_2} \sum_{f_j \in \mathcal{L}} V_2(f_i, f_j) p_j^t(f_j). \tag{28}$$

and the superscripts $t, t+1$ denote iteration numbers. The notation $j : (i, j)$ refers to a single sum over j such that (i, j) are pairs of sites belonging to a clique. Barring the different normalization techniques could be employed, Eqn. (28) is found in the update equations of various CRL algorithms [24, 11, 18]. There are, however, two differences. First, in most CRL problems, the first term of Eqn. (28) is absent and thus proper initialization of $\mathbf{p}$ is important. We will also omit this term from now on to emphasize the similarity with continuous relaxation labeling. Second, CRL problems typically involve maximization; thus, $V_2(f_i, f_j)$ would represent consistency as opposed to potential, and the update equation would add instead of subtract the gradient.

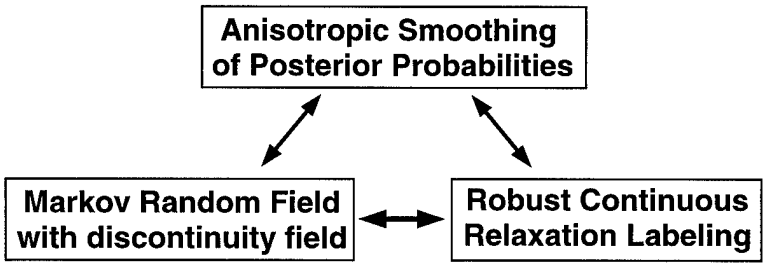


Figure 9: *Equivalence between anisotropic smoothing of posterior probabilities, Markov random fields with discontinuity fields, and robust continuous relaxation labeling.*

Isotropic Smoothing. A convenient way of visualizing the above operation is as isotropic smoothing. Since the sites represent pixels in an image, for each class f_i, $p_i(f_i)$ can be represented by an image (of posterior probabilities) such that k classes imply k such image planes. Together, these k planes form a volume of posterior probabilities. Each step of Eqn. (28) then essentially replaces the current estimate $p_i^t(f_i)$ with a weighted average of the neighboring assignment probabilities $p_j^t(f_j)$. In other words, the volume of posterior probabilities is linearly *filtered*. If the potential functions $V_2(f_i, f_j)$ favor similar labels, then the weighted average is essentially low-pass among sites with common labels and hi-pass among sites with differing labels.

15 Anisotropic smoothing

Isotropic smoothing causes significant blurring especially across region boundaries. A solution to this problem is to smooth adaptively such that smoothing is suspended across region boundaries and takes place only within region interiors. Anisotropic smoothing is often implemented by simulating nonlinear partial differential equations with the image as the initial condition [2, 22]. In this section, we show that while isotropic smoothing of posterior probabilities is the same as continuous relaxation labeling of a MRF, anisotropic smoothing of posterior probabilities is equivalent to continuous relaxation labeling of a MRF supplemented with a (hidden) analog discontinuity field. We also demonstrate that this method could also be understood as incorporating a robust consensus-taking scheme within the framework of continuous relaxation labeling. These relationships are depicted in Fig. 9.

We extend the original MRF problem to include a non-interacting, analog discontinuity field on a displaced lattice. Thus, the new energy to be minimized is:

$$E(\mathbf{f}, \mathbf{l}) = \sum_{(i,j)\in\mathcal{C}_2} \left[\frac{1}{2\sigma^2} V_2(f_i, f_j) \cdot l_{i,j} + (l_{i,j} - 1 - \log l_{i,j})\right] \tag{29}$$

where $V_1(f_i)$ has been dropped for simplicity since the discontinuity field does not interact with it. The individual sites in the discontinuity field $\mathbf{l}$ are denoted by $l_{i,j}$ which represent either the horizontal or vertical separation between sites i and j in $\mathcal{S}$. When $l_{i,j}$ is small, indicating the presence of a discontinuity, the effect of the potential $V_2(f_i, f_j)$ is suspended; meanwhile, the energy is penalized by the second term in Eqn. (29). There are a variety of

penalty functions that could be derived from the robust estimation framework (see Black [3]). The penalty function in Eqn. (29) was derived from the Lorentzian robust estimator.

The minimization of $E(\mathbf{f}, \mathbf{l})$ is now over both $\mathbf{f}$ and $\mathbf{l}$. Since the discontinuity field is non-interacting, $\mathbf{l}$ can be minimized analytically by computing the partial derivatives of $E(\mathbf{f}, \mathbf{l})$ with respect to $l_{i,j}$ and setting that to zero. Doing so and inserting the result back into $E(\mathbf{f}, \mathbf{l})$ gives us

$$E(\mathbf{f}) = \sum_{(i,j)\in\mathcal{C}_2} \log\left[1 + \frac{1}{2\sigma^2} V_2(f_i, f_j)\right]. \tag{30}$$

Rewriting this equation in a form suitable for CRL, we get

$$E(\mathbf{p}) = \sum_{(i,j)\in\mathcal{C}_2} \log\left[1 + \frac{1}{2\sigma^2} \sum_{(f_i,f_j)\in\mathcal{L}^2} V_2(f_i, f_j) p_i(f_i) p_j(f_j)\right]. \tag{31}$$

Note that when $p_i(f_i)$ is restricted to $\{0, 1\}$, these two equations are equivalent.

Anisotropic Smoothing. To compute the update equation for CRL, we take the derivative of $E(\mathbf{p})$ with respect to $p_i(f_i)$:

$$\frac{\partial E(\mathbf{p})}{\partial p_i(f_i)} \doteq \sum_{j:(i,j)\in\mathcal{C}_2} w_{i,j} \left[\sum_{f_j\in\mathcal{L}} V_2(f_i, f_j) p_j(f_j)\right] \tag{32}$$

where

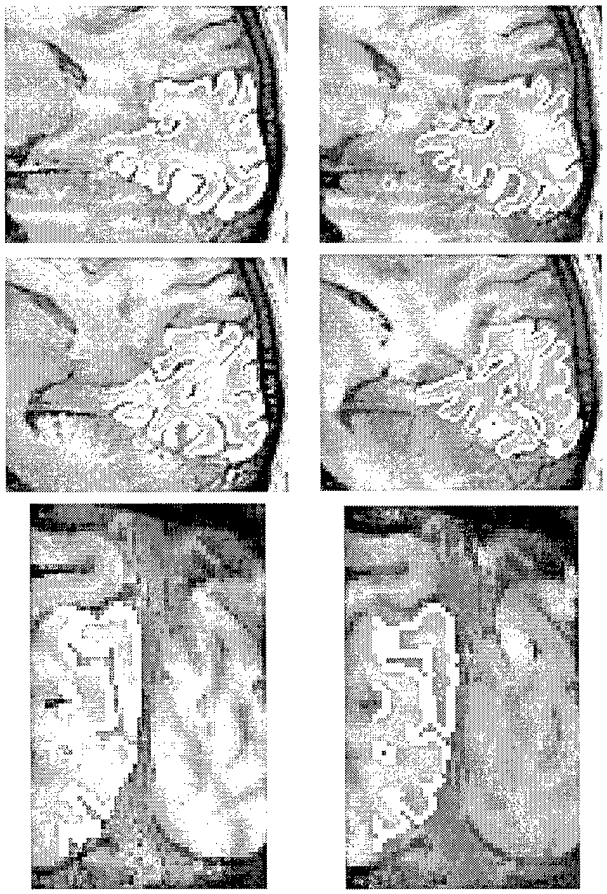
$$w_{i,j} = 2\sigma^2 / \left[2\sigma^2 + \sum_{(f_i,f_j)\in\mathcal{L}^2} V_2(f_i, f_j) p_i(f_i) p_j(f_j)\right]. \tag{33}$$

The term $w_{i,j}$ encodes the presence of a discontinuity. If $w_{i,j}$ is constant, then the above equation reverts to the isotropic case. Otherwise, $w_{i,j}$ either enables or disables the penalty function $V_2(f_i, f_j)$. This equation is similar to the anisotropic diffusion equation proposed by Perona and Malik [22].

Robust Continuous Relaxation Labeling. Each iteration of continuous relaxation labeling can be viewed as a consensus-taking process [33]. Neighboring pixels vote on the classification of a central pixel based on their current assignment probabilities $p_j(f_j)$, and their votes are tallied using a weighted sum. The weights used are the same throughout the image; thus, pixels on one side of a region boundary may erroneously vote for pixels on the other side. Anisotropic smoothing of the posterior probabilities can be regarded as implementing a robust voting scheme since votes are tempered by $w_{i,j}$ which estimates the presence of a discontinuity. The connection between anisotropic diffusion on continuous-valued images and robust estimation was recently demonstrated by Black *et. al.* [4]. We plan to further investigate the use of results in [4] for posterior diffusion.

16 Results and discussion

The anisotropic smoothing scheme was used to segment white matter from MRI data of human cortex. Pixels at a given distance from the boundaries of the white matter classification were then automatically classified as gray matter. Thus, gray matter segmentation relied heavily on the white matter segmentation being accurate. Fig. 10 shows comparisons between gray matter segmentations produced automatically by the proposed method and those obtained manually. More examples can be found in [32].



(a) (b)

Figure 10: *Left images show manual gray matter segmentation results; right images show the automatically computed gray matter segmentation.*

The technique being proposed bears some superficial resemblance to schemes that anisotropically smooth the raw image before classification [14]. Besides the connection between our technique and MAP estimation of Markov random fields, which is absent in schemes that smooth the image directly, there are two other important differences. First, anisotropic smoothing of the raw image does not take into consideration the discrete number of classes that are actually present. Second, anisotropic smoothing of the raw image is only applicable when the noise corrupting the image is additive and class independent. For example, if the class means were identical and the classes differed only in their variances, then anisotropic

smoothing of the raw image is ineffective. On the other hand, applying anisotropic smoothing on the posterior probabilities is still feasible even when the class likelihoods are described by general probability mass functions.

The equivalence between anisotropic smoothing of posterior probabilities and MRF's with discontinuity fields also offers a solution to the problems of edge handling and missing data. These two issues can be treated in the same manner as in traditional regularization. Solving of the latter implies that MAP classification can be obtained even at locations where the pixel values are not provided.

References

[1] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel, "Axioms and fundamental equations of image processing," *Arch. Rational Mechanics and Anal.* **16:IX**, pp. 200-257, 1993.

[2] L. Alvarez, P. L. Lions, and J. M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM-JNA* **29**, pp. 845-866, 1992.

[3] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int'l J. Computer Vision*, 19(1):57–91, 1996.

[4] M. Black, G. Sapiro, D. Marimont, D. Heeger, "Robust anisotropic diffusion," *IEEE-IP*, to appear.

[5] P. Blomgren and T. Chan, "Color TV: Total variation methods for restoration of vector valued images," *UCLA TR*, 1996.

[6] V. Caselles, J-L. Lisani, J-M. Morel, and G. Sapiro, "Shape preserving local histogram modification," *HP Labs Technical Report* **97:58**, April 1997.

[7] K. Castleman, *Digital Image Processing*, Prentice-Hall, 1996.

[8] T. Chan and P. Mulet, "Iterative methods for total variation image restoration," *UCLA TR* **96:38**, October 1996.

[9] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE-IP*, to appear.

[10] R. Cromartie and S. M. Pizer, "Edge-affected context for adaptive contrast enhancement," *Proceedings Information Processing in Medical Imaging, Lecture Notes in Comp. Science* **511**, pp. 474-485, Wye-UK, July 1991.

[11] O. D. Faugeras and M. Berthod, "Improving consistency and reducing ambiguity in stochastic labeling: an optimization approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 3:412–423, 1981.

[12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(6):721–742, 1984.

[13] D. Geman and C. Yang, *IEEE-IP* 4:7, pp. 932-946, 1995.

[14] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Trans. Medical Imaging*, 11:221–232, 1992.

[15] R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, 1987.

[16] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions.* John Wiley & Sons, New York, 1986.

[17] P. J. Huber. *Robust Statistics.* John Wiley & Sons, New York, 1981.

[18] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5(2):267–286, 1983.

[19] N. Nordström, "Biased anisotropic diffusion: A unified regularization and diffusion approach to edge detection," *Image and Vision Computing* **8**, pp. 318-327, 1990.

[20] S. Z. Li, H. Wang, and M. Petrou, "Relaxation labeling of Markov random fields," In *Int'l Conf. Pattern Recognition*, pages 488–492, 1994.

[21] S. J. Osher and J. A. Sethian, "Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics* **79**, pp. 12-49, 1988.

[22] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE-PAMI* **12**, pp. 629-639, 1990.

[23] B. Romeny (Ed.). *Geometry Driven diffusion in Computer Vision*, Kluwer, 1994.

[24] A. Rosenfeld, R. Hummel, and S. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Systems, Man, and Cybernetics*, 6(6):420–433, 1976.

[25] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, 1987.

[26] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, pp. 259-268, 1992.

[27] J. Serra, *Image Analysis and Mathematical Morphology, vol. 2: Theoretical Advances.*, Academic Press, 1988.

[28] G. Sapiro and V. Caselles, *Histogram Modification via PDE's, Journal of Differential Equations* **135:2**, pp. 238-268, 1997.

[29] G. Sapiro and D. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE-IP* **5**, pp. 1582-1586, 1996.

[30] J. A. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Sciences,* Cambridge University Press, Cambridge-UK, 1996.

[31] L. Schwartz, *Analyse I. Theorie des Ensembles et Topologie.*, Hermann, 1991.

[32] P. C. Teo, G. Sapiro, and B. A. Wandell, "A method of creating connected representations of cortical gray matter for functional MRI visualization," *IEEE Trans. on Medical Imaging*, submitted, 1997.

[33] Y. Weiss and E. Adelson, "Perceptually organized EM: a framework for motion segmentation that combines information about form and motion," In *Int'l Conf. Computer Vision and Pattern Recognition*, pages 312–326, 1996.

[34] Y. L. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE-IP* **5**, pp. 1539-1553, 1996.

Part V
Papers under ONR-N00014-97-1-0509

1. "Robust anisotropic diffusion and sharpening of scalar and vector images," *Proc. IEEE-International Conference on Image Processing*, Santa Barbara, California, October 1997 (with M. Black, D. Marimont, and D. Heeger).

2. "Shape preserving local contrast enhancement," *Proc. IEEE-International Conference on Image Processing*, Santa Barbara, California, October 1997 (with V. Caselles, J-L. Lisani, and J-M. Morel).

3. "Anisotropic diffusion of posterior probabilities," *Proc. IEEE-International Conference on Image Processing*, Santa Barbara, California, October 1997 (with P. Teo and B. Wandell).

4. "Color and illuminant voting," *Proc. International Conference on Computer Vision*, Bombay, India, January 1998, to appear.

5. "Affine invariant symmetry sets and skew symmetry," *Proc. International Conference on Computer Vision*, Bombay, India, January 1998, to appear (with P. Giblin).

6. "Segmenting cortical gray matter for functional MRI visualization," *Proc. International Conference on Computer Vision*, Bombay, India, January 1998, to appear (with P. Teo and B. Wandell).

7. "Color snakes," *Computer Vision and Image Understanding*, to appear.

8. "Robust anisotropic diffusion," *IEEE Trans. Image Processing*, to appear (with M. Black, D. Marimont, and D. Heeger).

9. "Contrast enhancement via image evolution flows," *Graphical Models and Image Processing*, to appear (with V. Caselles).

10. "Shape preserving local contrast enhancement," *HP Labs Technical Report* **97:58**, April 1997 (with V. Caselles, J-L. Lisani, and J-M. Morel).

11. "Affine invariant evolution of non-convex curves," submitted (with S. Angenent and A. Tannenbaum).

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	10/20/97	5/1/97 -- 10/31/97

4. TITLE AND SUBTITLE

Structure and Geometry in Partial Differential Eqn. in Image Processing and Analysis

5. FUNDING NUMBERS

ONR-N00014-97-1-0509

6. AUTHOR(S)

Guillermo Sapiro

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Minnesota, Department of ECE
Minneapolis, MN 55455

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the ONR **position, policy, or decision, unless so designated by other documentation.**

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We first describe results on shape-base image contrast enhancement. This work shows how to perform local contrast enhancement while preserving the shapes in the image. We have transfered this software to Dr. Szymczak from Physical Acoustics at the Naval Research Laboratory for testing on underwater laser images (LIDAR). We then show a novel approach to anisotropic diffusion. This approach is based on robust statistics, and in particular, in the theory of influence functions. This technique is, for example, of particular significance for image denoising prior to segmentation. We conclude this document with a description of a novel technique of incorporating prior information in anisotropic diffusion. The idea is to use Bayes rule to compute posterior distributions, and then, regularize those distributions via partial differential equations before the MAP is computed.

Although a number of results have already been obtained in these areas, the work described in this document is still in progress. As was mentioned above, we want to extend the work on shape-preserving contrast enhancement to include additional definitions of shape and adapt it to specific applications. We are also planing to extend the robust framework for anisotropic diffusion to vector-valued data, and investigate fast implementations. We plan to further investigate the underlying theory of the posterior diffusion work, and to apply it to additional problems. The work described in this document opens a number of theoretical questions that we plan to address as well.

14. SUBJECT TERMS

Image processing and analysis, PDE's, contrast enhancement, anisotropic diffusion, Bayes ryle.

15. NUMBER OF PAGES

28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102